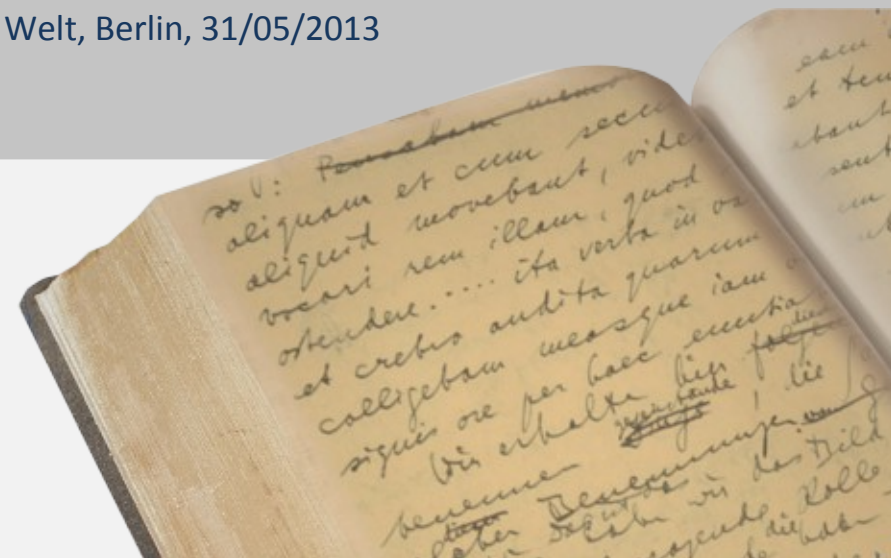


Semantische Nachhaltigkeit und Kontrolle: Gedanken zu schema.org und Linked Open Data

Prof. Dr. Stefan Gradmann (KU Leuven)

Nachhaltigkeit in der digitalen Welt, Berlin, 31/05/2013



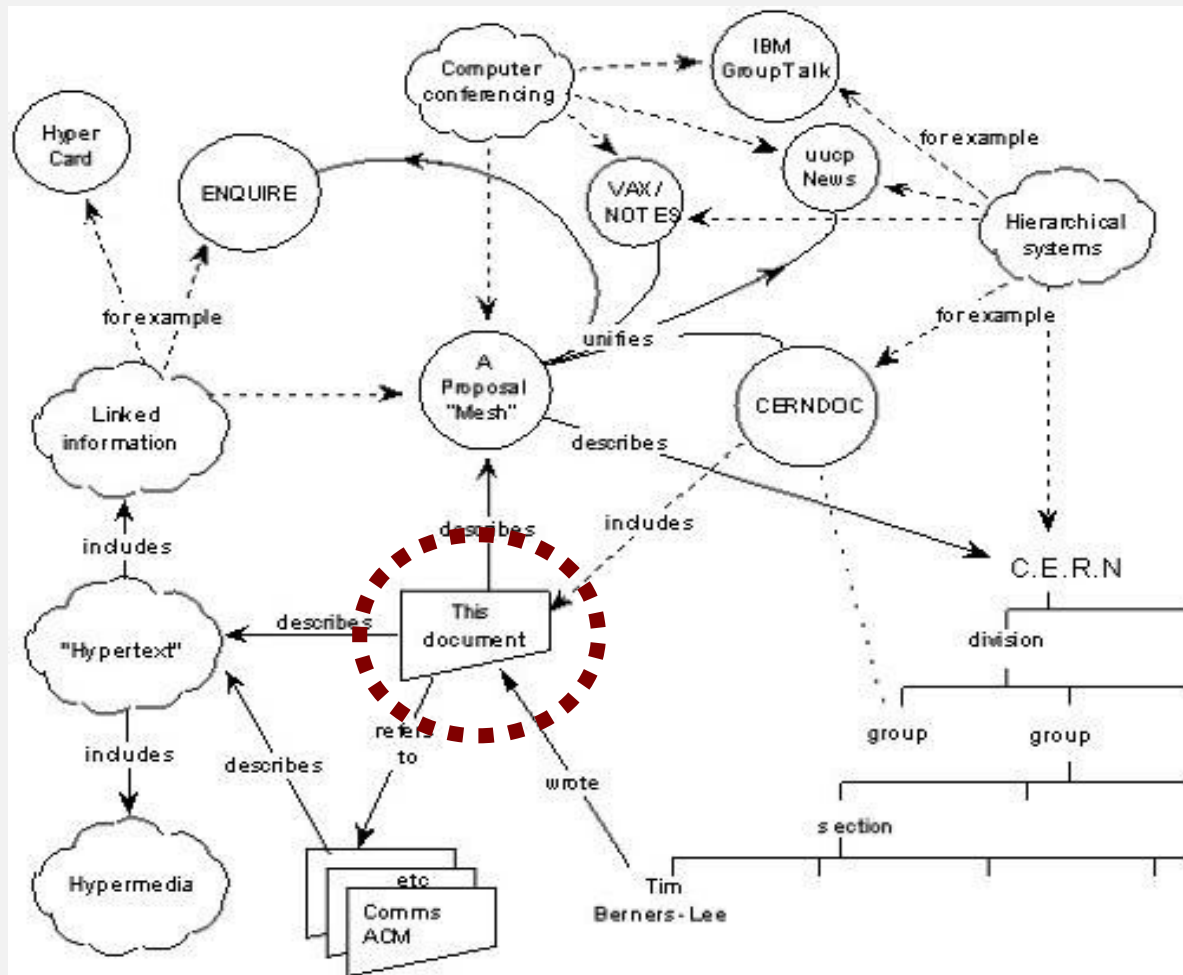
- Linked Open Data: ein fröhliches Chaos
- Schema.org: Ordnung, Disziplin und Kontrolle
- Eine versteckte Agenda?
- Semantischer Darwinismus oder Zensur??

Linked Open Data: ein fröhliches Chaos

Das WWW: ein fröhliches, robustes Chaos

“It's massively successful. It is trivially simple.
Massively successful like karaoke - anybody can do it.”
(Ted Nelson 2001)

“Anyone can say anything about anything”
(Tim Berners Lee, 2002)

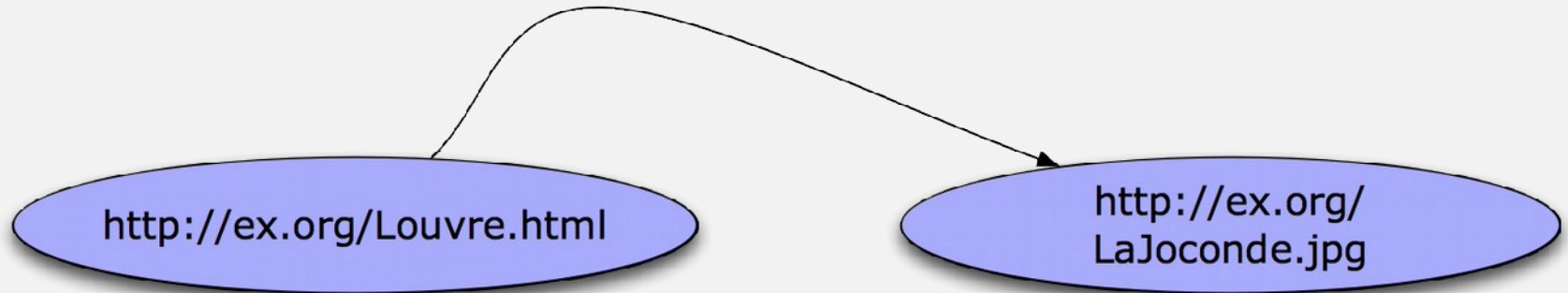


Information Management:
A Proposal
(TBL, 1989)

... zweifach
erweitert:

- syntaktisch
- im Repräsentationsbereich

Entitäten und Links im Web der Dokumente

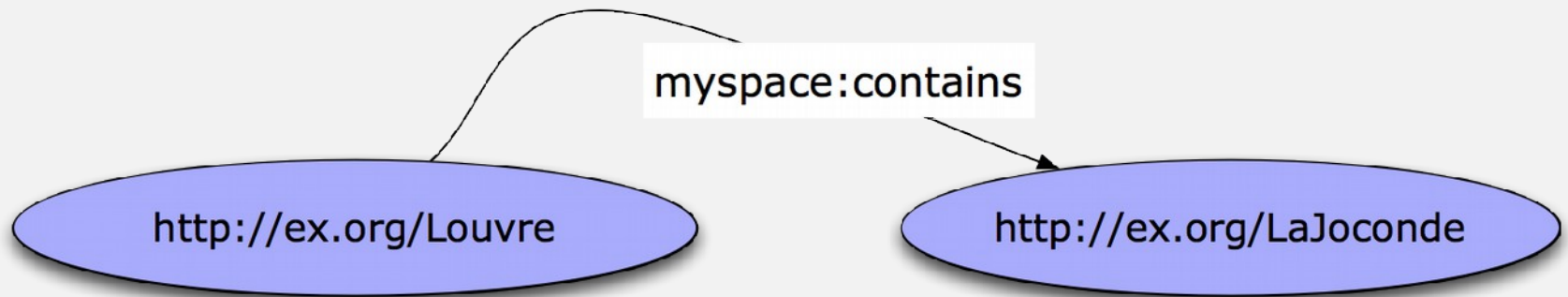


```
<a href="http://ex.org/LaJoconde.jpg">b</a>
```

- Entitäten (resources) haben HTTP URIs zur Identifikation und Links, die sie verbinden – aber es fehlt doch so einiges!
- Was für eine Art von Entitäten sind 'Louvre.html' und 'LaJoconde.jpg'?
 - Die Maschine weiß es nicht und kann es auch nicht im Ansatz erraten.
 - Menschen wissen es vielleicht auch nicht, erkennen aber implizierten Kontext!
- Welcher Art ist die Verbindung zwischen ihnen?
 - Die Maschine weiß es nicht und kann es auch nicht im Ansatz erraten.
 - Menschen wissen es vielleicht auch nicht, erkennen aber implizierten Kontext!

Syntaktische Erweiterung des Web der Dokumente: RDF (1)

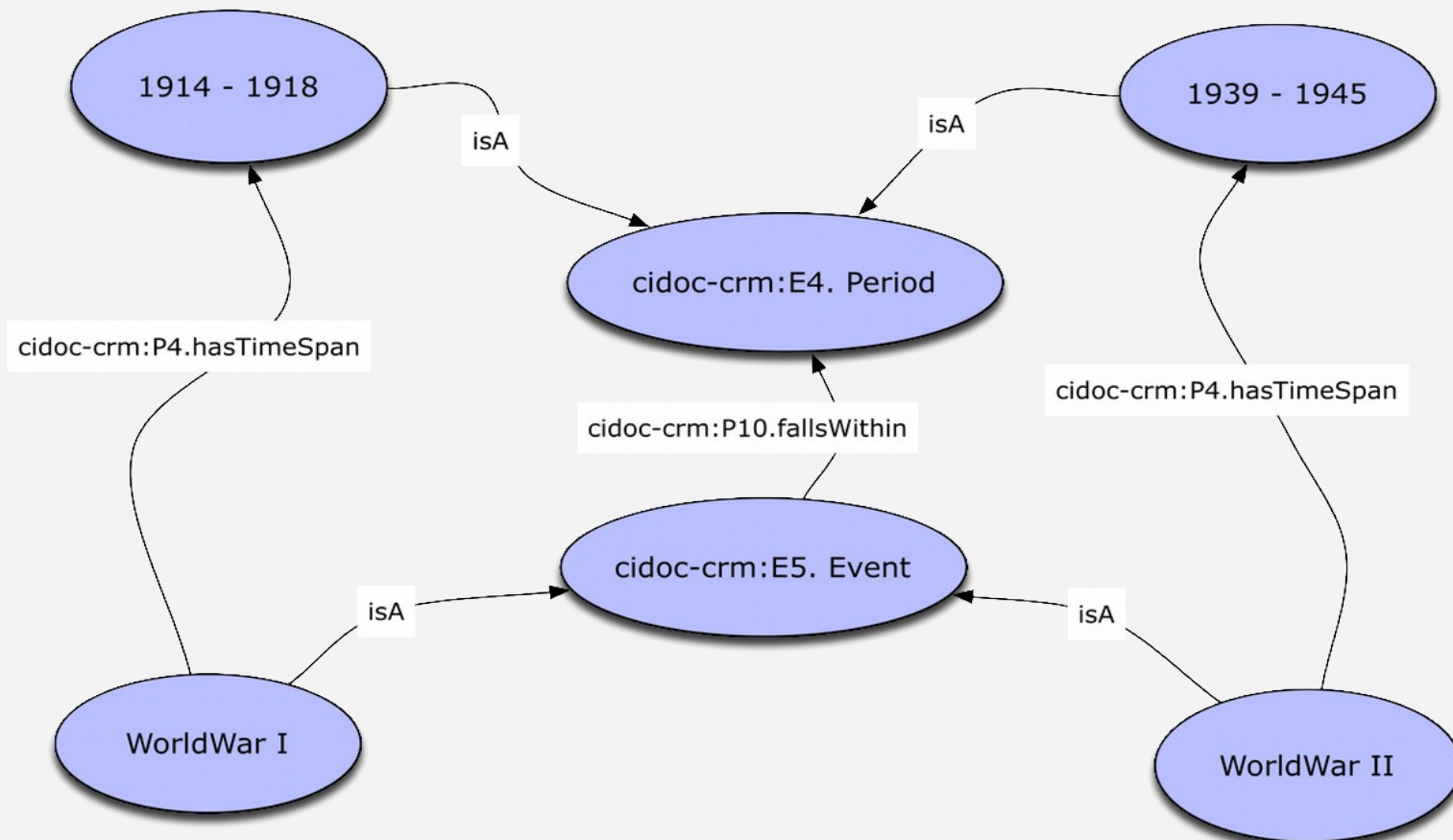
- Spezifikation einer Syntax für Aussagen über Entitäten im WWW: Resource Description Framework (RDF) Tripel



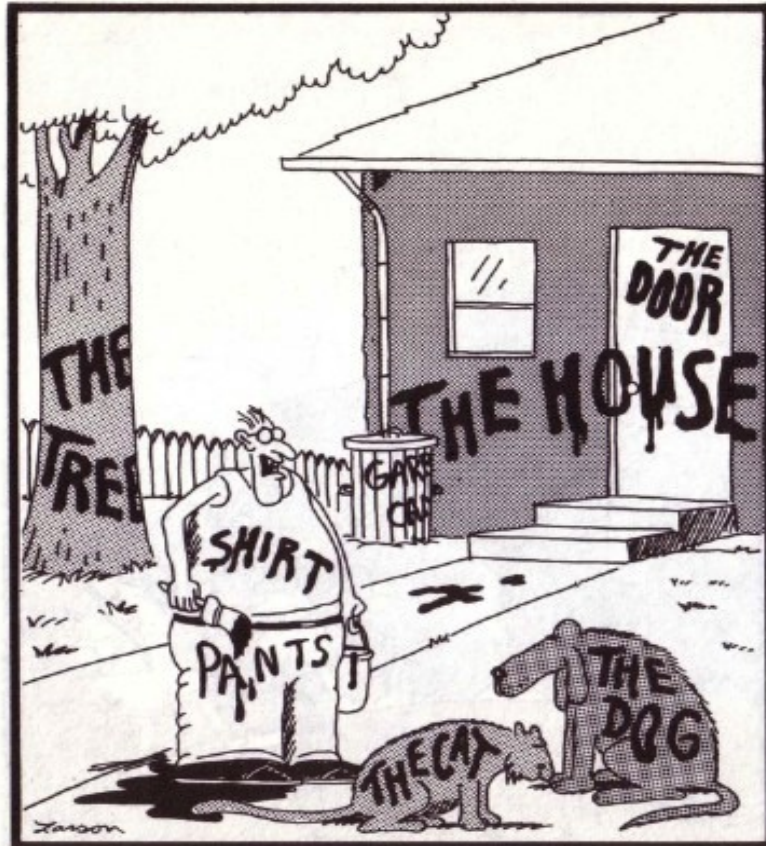
- Spezifikation einer Schema-Sprache (RDFS) als 'Grammatik' mit Elementen wie
 - Klassen ('Stuhl' als Instanz der Klasse Stuhl), Eigenschaften, Werten
 - Hierarchien von Klassen und Eigenschaften ('Stühle' als Unter-Klasse von 'Möbel', 'lehrt' als Unter-Eigenschaft von 'kommuniziert')
 - Vererbung ('kommunizieren' basiert auf Sprache → 'lehren' ebenso)
 - Unterstützung für einfache deterministische logische Operationen

Syntaktische Erweiterung des Web der Dokumente: RDF (2)

- So können wir Strukturen in Aggregationen von Tripeln etablieren: das Resultat sind leichgewichtige 'Ontologien':



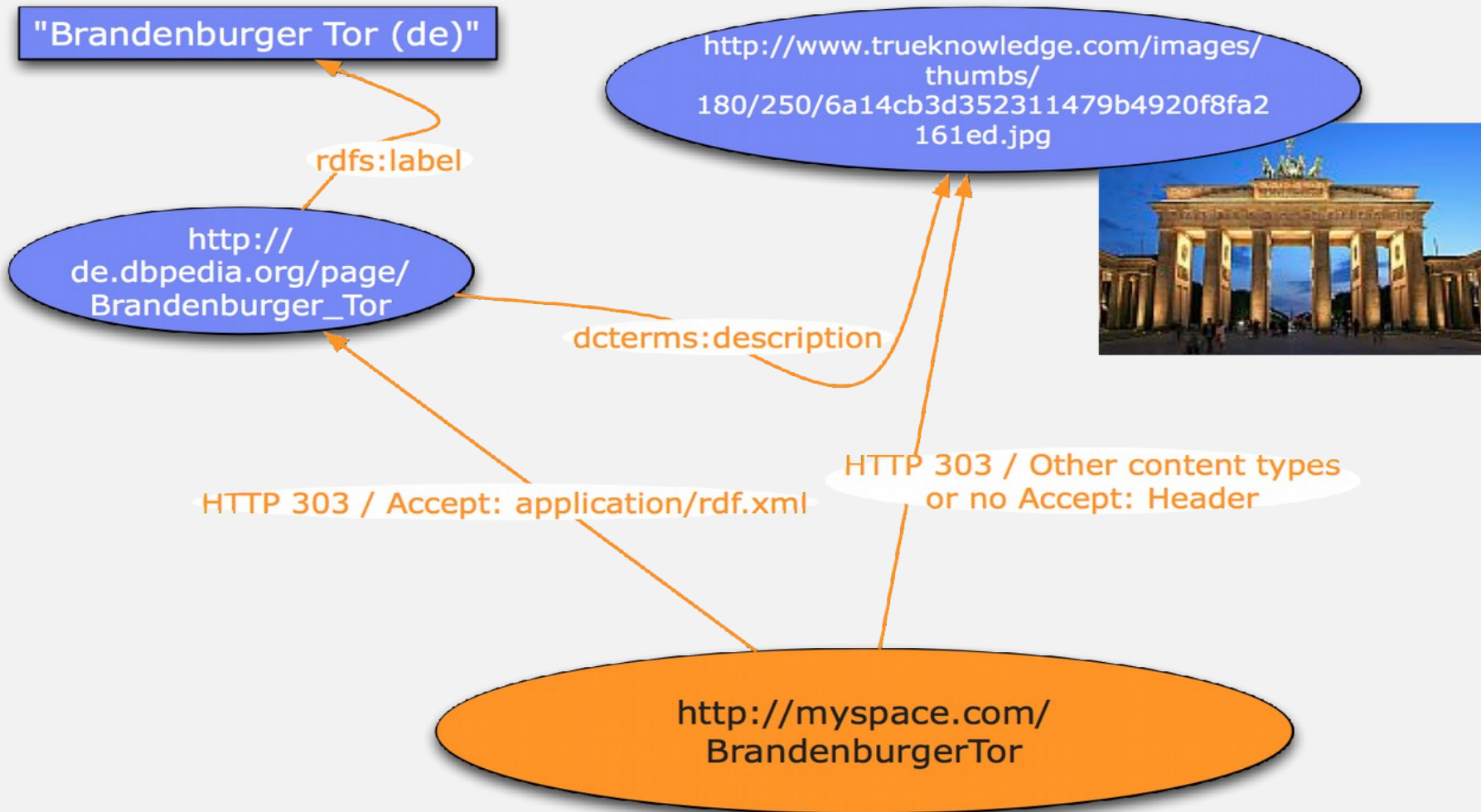
Erweiterung des Repräsentationsbereichs: Das Web der Dinge ... nicht ganz so:



Was stimmt
an diesem Bild **nicht**?

Zitiert nach Ronald Carpentier's
Blog unter
<http://carpentier.wordpress.com/2007/08/08/1-2-3/>

Erweiterung des Repräsentationsbereichs zu einem 'Web der Dinge'

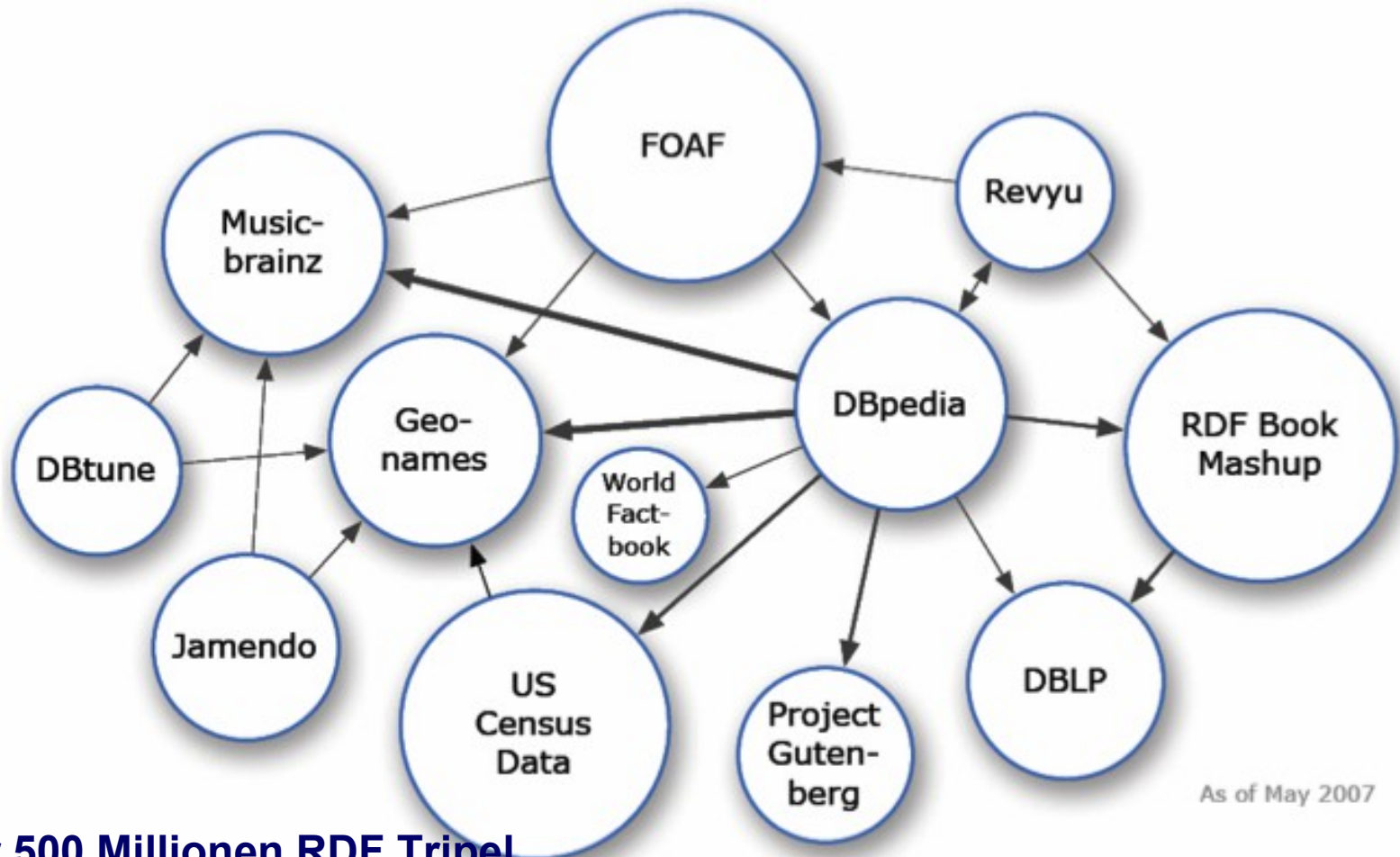


Linked Data essentials

1. Use URIs Standard-**Identifizier**
2. Use HTTP URIs Standard-**Zeiger**
3. Serve useful information using SPARQL, RDF standards
4. Mention URIs of related objects Standards für **Fragen**
und **Aussagen**
Links zu Kontextentitäten

Copyright © 2008 W3C (MIT, ERCIM, Keio)

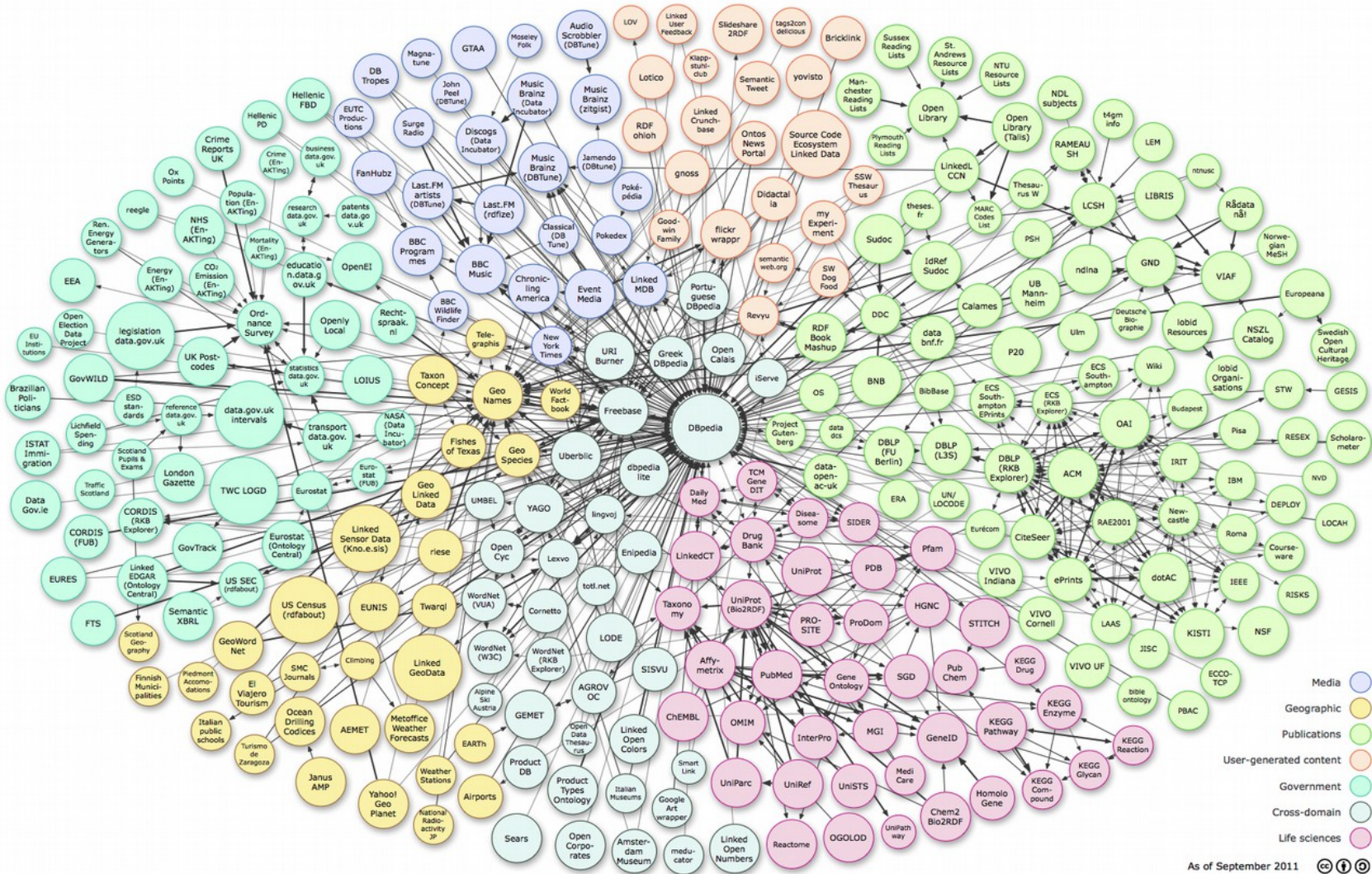
[http://www.w3.org/2008/Talks/0617-lod-tbl/#\(4\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(4))



- Über 500 Millionen RDF Tripel
- Ca. 120.000 RDF-Links zwischen Datenquellen

© Richard Cyganiak

Und eine Menge Kugeln Ende 2011

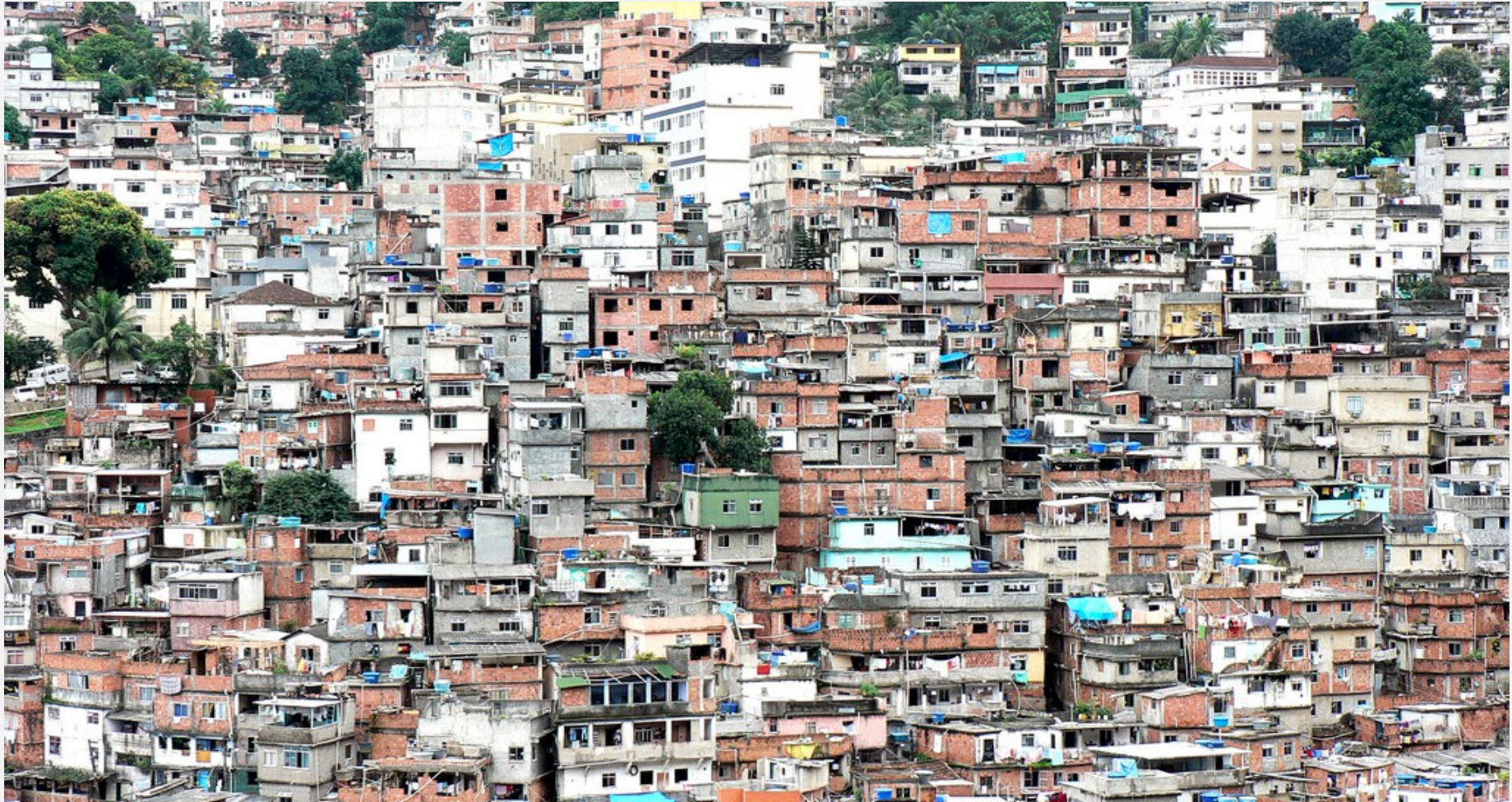


Andere Darstellungsformen von LoD – und ein Problem

- Klassen:
 - <http://inkdroid.org/lod-graph/>
- Instanzen:
 - <http://zoom.it/Vj6F> or
<http://bgriffen.scripts.mit.edu/www/media/json/thinkers/>
- Vokabulare:
 - <http://lov.okfn.org/dataset/lov/>
- Personen ...:
 - <http://ws.nju.edu.cn/falcons/ontologysearch/result.jsp?query=person>

- To facilitate operation at Internet scale, RDF is an **open-world framework** that allows **anyone to say anything about anything**. In general, it is not assumed that all information about any topic is available. A consequence of this is that RDF cannot prevent anyone from making **nonsensical** or **inconsistent** assertions, and applications that build upon RDF must find ways to deal with **conflicting sources of information**. (This is where RDF departs from the XML approach to data representation, which is generally quite prescriptive and aims to present an application with information that is well-formed and complete for the application's needs.) (W3C, <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/>)
- **Issues:** Versioning, provenance, integrity ... and **(partial) semantic redundancy!**

LOD-Architektur: die Cloud als Favela



Warum nicht so?



Kathedralen bauen



Eine fertige Kathedrale



Schema.org: Ordnung, Disziplin und Kontrolle

- Lanciert am 02. Juni 2011
- Von Google, Bing, Yahoo. Yandex kam später dazu
- Ziel: “create and support a standard set of schemas for structured data markup on web pages”
(<http://blog.schema.org/search?updated-max=2011-12-12T12:10:00-08:00>)
- Kontrolliertes Kernvokabular und Erweiterungen
- Ursprünglich auf microdata beschränkt, RDFa kam im September 2011 dazu
- Die Welt in schema.org: <http://schema.org/docs/full.html>
- Personen in schema.org: <http://schema.org/Person>

- Generell interessant als Quelle:
<http://linter.structured-data.org/examples/>
- Filme
 - <http://www.imdb.com/title/tt0306414/>
 - http://www.rottentomatoes.com/m/the_adventures_of_tin_tin/

Eine versteckte Agenda?

Darin Stewart:

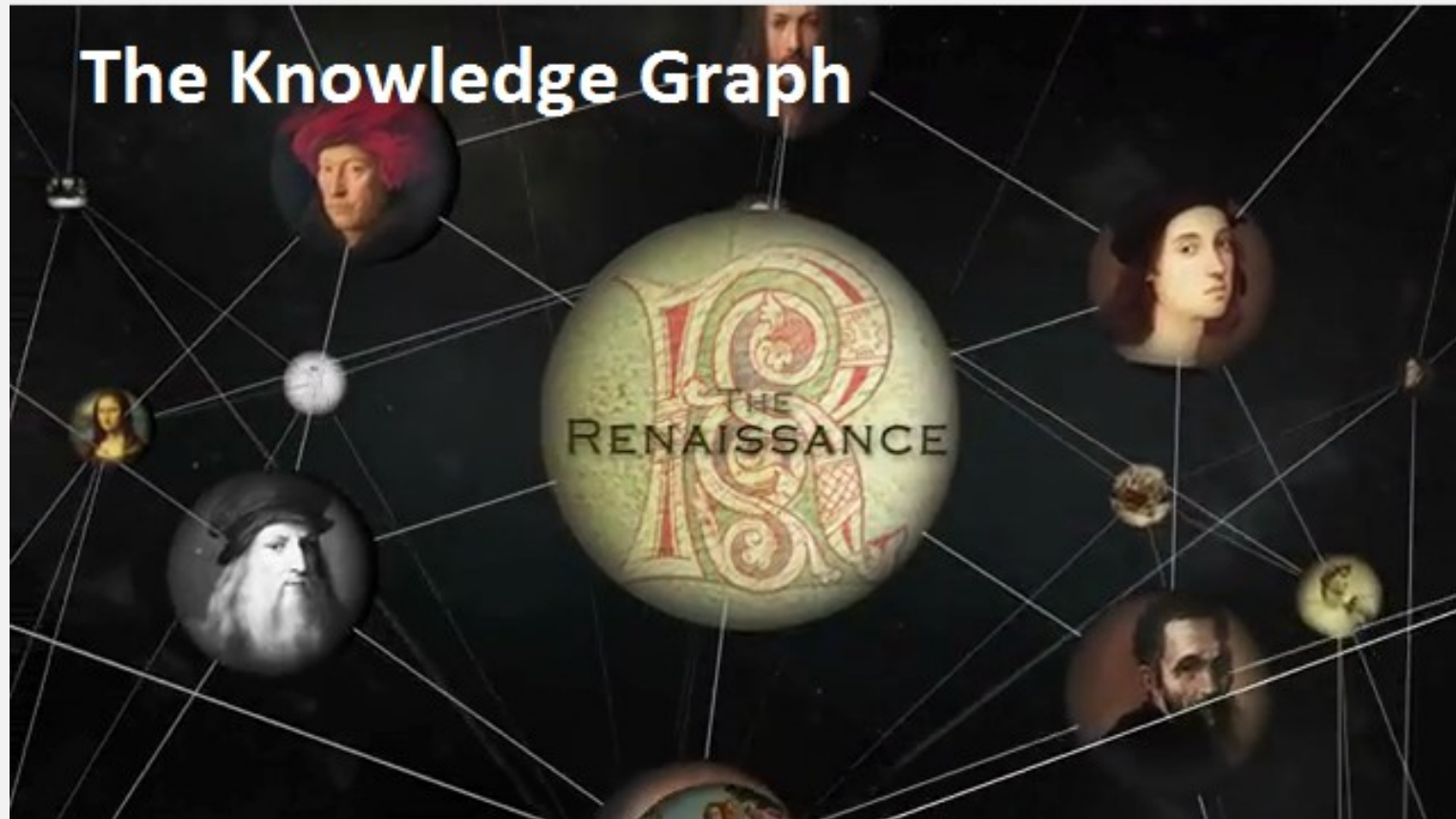
- “Schema.org appears to be Linked Data Lite with extremely limited support for vocabularies outside of the service.”
- “There is a subtle air of intimidation throughout the schema.org announcements and documentation.”
- “Again, I could just be paranoid, but this is Microsoft and Google we’re talking about. Whatever happened to “do no evil?””

(all from <http://blogs.gartner.com/darin-stewart/2011/06/04/schema-org-webmaster-one-stop-or-linked-data-land-grab/>)

Schema.org: wo ist die Motivation?

- Google, Bing, Yahoo und Yandex sind harte Konkurrenten.
- Was bringt sie dazu, in einem Kern-Geschäftsbereich zu kooperieren?
- Warum verbiegen sie Standards (und sei es auch nur ein bisschen)?
- Ist schema.org eine Reaktion auf Linked open Data??
- Sind wir zurück bei The Cathedral vs. The Bazaar again (Eric Raymond!) ???

Und wie hängt das alles zusammen mit ...



=> <http://manu.sporny.org/2012/google-indexing-schema-rdfa/>

Semantischer Darwinismus oder Zensur??

- Nein:
 - Niemand hat Google in diesem Sinne beauftragt.
 - Schema.org ist nur eine radikale Variante dessen was wir bald 'ontologischen Darwinismus' nennen könnten.
- Ja:
 - Faktisch handelt es sich letztendes um Zensur.
 - Niemand hat dies in Auftrag gegeben: ein 'outsourcing' findet nicht statt.
 - Die Wirkungen sind aber ähnlich, wenn wir Google gewähren lassen.
 - Und sie passen in Google's Geschäftsmodell ...

Vielleicht ist es schlimmer als Zensur?

- Traditionelle Zensur war der Versuch, Menschen an der **Veröffentlichung** ihres Schaffens zu hindern.
- Die Sprache und künstlerischen Hilfsmittel waren dabei **niemandes Eigentum**.
- Mit schema.org werden die Ausdrucksmittel selbst **oligopolistisches Eigentum**.
- ... bei schema.org geht es darum, **was gesagt werden kann** (bzw. nicht gesagt werden kann) und welche Aussagen effektiv **wahrgenommen** werden (und welche einfach nicht registriert werden).
- Bei schema.org geht es nicht um Geld, sondern um die Leitwährung des WWW: **Aufmerksamkeit!**

Fragen?